DOCUMENT RESUME

ED 124 600 TM 005 360

AUTHOR Law, Alexander I.
TITLE State Level Evaluation Operations in a Multifunded Program Context.
PUB DATE [Apr 76]
NOTE 5p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Educational Finance; *Evaluation Needs; *Federal Programs; Program Effectiveness; *Program Evaluation; Research Problems; School Districts; *State Programs
IDENTIFIERS California

ABSTRACT

Several years ago, California, along with other states, discovered that there was a layering effect of categorical programs, both state and federal. Most programs were aimed at unique populations--disadvantaged, bilingual, etc. This proliferation and the evaluation requirements inherent in each of the categorical efforts created problems of sufficient magnitude to warrant action to relieve school districts. Through reform efforts schools were required to provide thorough needs assessments of their pupils to develop plans in which the various funding sources could be brought together into a coherent whole to meet the established needs of the students. While this reform movement has forged ahead the problem remains that there are still unique evaluation requirements for each of the individual programs. The problem arises of attempting to make reasonable judgments about the total effectiveness of a reform effort and yet having to meet the legislatively established requirements of unique funding source evaluations. This dilemma presents itself: The kind of information policy makers need is not of the descriptive nature, which has typically characterized an evaluation, but rather one which can yield inferences about contrasts between programs. A movement toward a "hyper-evaluation" which is more akin to experimental or research design is foreseen. (RC)

1307

# STATE LEVEL EVALUATION OPERATIONS IN A MULTIFUNDED PROGRAM CONTEXT*

Alexander I. Law
California State Department of Education

The subtitle of this paper might logically be, "Hard Times Lead to Hard Decisions." You are all familiar with the evolution in the past decade of the evaluation and accountability movement which was generated primarily by passage of the elementary and secondary education programs. In addition to these ongoing requirements of reporting to the federal government, in California we have had certain developments which have amplified the problem.

These developments can be called reform efforts. Several years ago, we, along with other states, discovered that there was a layering effect of categorical programs, both state and federal. The majority of these programs were aimed at unique populations, such as the disadvantaged, Bilingual--students with unique educational needs. This proliferation and the evaluation requirements inherent in each of the categorical efforts created problems of sufficient magnitude to warrant action to relieve school districts.

Among these problems were the proliferation of paper work and the need to assess multiple programs that were dealing with the same pupil populations. An extreme example is that in a single second grade classroom in one major metropolitan area, there were seven separately funded programs, each with unique application, program development and evaluation requirements. The absurdity of such a situation is self-evident. Three years ago a group of prominent educators conceptualized the idea of a massive reform, beginning in the elementary grades. This effort has been known as the Early Childhood Education Reform (ECE). I purposely use the word 'reform' rather than program because ECE attempts to make substantial changes in the total education program by not only instituting changes in instructional practices, but also addressing the problem of fragmented efforts to assist students. Under this reform concept schools are required to provide thorough needs assessments of their pupil populations to develop plans in which the various funding sources can be brought together into a coherent whole to meet the established needs of

*A presentation in a symposium, An End of Affluence: Educational Evaluation in Tight Money Times, Annual Meeting of the American Educational Research Association, San Francisco, California, April 19-23, 1976.

their students. Thus, within ECE we see the combination of Title I, Bilingual education, both federal and state, our own Educationally Disadvantaged Youth program, special reading programs, and certain others all brought together in a unified effort. While this movement has forged ahead, and gained popularity from a practical view, the problem remains that there are still unique evaluation requirements for each of the individual programs.

We find ourselves cast into the situation of attempting to make reasonable judgments about the total effectiveness of a reform effort and yet having to meet the legislatively established requirements of unique funding source evaluations.

This effort rapidly becomes nonsensical when you consider two evident facts (1) the funding sources do not define as such a unique instructional program They are merely vehicles by which dollars are transferred from one treasury to another, and (2) the populations for whom the programs are designed are not uniquely different in the various schools in which the programs are implemented. Because of the hard money times, we are asked now to make interprogram comparisons--that is, to contrast effects of one categorical program with those of another, when in fact they are impacting the same populations and most probably using the same instructional interventions. Demands are made on us by the legislature and regulatory and control agencies in the name of sound public policy, to determine the relative worth of the comparable programs. Is the Miller-Unruh Reading Program more effective or less effective than a reading program funded by the state program for disadvantaged youth?

Under the historical mode of federally funded categorical programs, dollars were apportioned on a formulaic basis. They were entitlement programs. Districts were entitled to receive money, and it still virtually takes a felonious act on the part of a district not to get these dollars. No judgment of relative success or failure is really necessary for these programs to be continued. The state categorical programs however, are based on a contrary view. That is, these programs must demonstrate their effectiveness if they are to be continued. Thus, in addition to the between or interprogram comparisons, we now must make some judgment about the relative effectiveness of schools. Such a process is obviously frought with peril. Policy makers assume that we are able to make a true evaluation--a completely accurate ranking of effectiveness from the greatest to the least--regardless of the pupil population, the instruments involved, and all the intervening variables.

The situation is not unlike that which occurs with regard to interpretation of grade equivalent scores, when people know that John's grade equivalent of 5.6 is obviously superior to Jane's equivalent of 5.5.

This misunderstanding leads to the last of my litany of problems, of appropriate instrumentation and analysis for use in evaluating and reporting results. Historically, we have required a pre-post procedure using norm-referenced tests. We are able to take these various test results and put them on a common metric such as a standard score, which I believe is infinitely more sensible than other scales. However, the problem of test content, the technical problems of norms, and a myriad of others remain to be contended with. What have we done to attempt to resolve these dilemmas? The first move was to go to a consolidated application evaluation and applicant agencies could in one document apply for all categorical funds. We moved to a consolidated evaluation where in one document districts could report to us data necessary for analysis. We issued guidelines to districts where one assessment could be used for a variety of categorical programs. We are contemplating and weighing the merits of using our state assessment program as the prime vehicle to collect common achievement information in the elementary grades. These provide partial solutions to the problem but the tough questions have yet to be answered.

We seek the advice of various groups to help us in the problem of analysis and the presentation of appropriate information to the body politic. I use the phrase "body politic" in the broadest sense, since it is increasingly clear that we are becoming cast into a political mode. Every evaluation report becomes a political document, and its worth is cherished by some and ridiculed by others. The evaluation reports are not used as the sole input by which broad policy decisions are made. Evaluation reports are not based upon rigorous experimental designs, yet inferential statements are demonstrated. These self-evident facts are difficult to communicate. The evaluation community spends an inordinate amount of time qualifying and issuing caveats about evaluation reporting. The net effect is to reduce the credibility of evaluation reports.

Somehow the credibility of evaluation must be re-established. Academic nit-picking as to the relative superiority of analytical methods does not assist any agency in attempting to communicate information to policy makers.

Evaluators cannot agree on a design, the appropriate instrumentation, or the appropriate analytical procedure. Phillips, in his paper, "When Evaluators Disagree," states the problem well. His solution is to establish panels of experts with common philosophical backgrounds. I'm not sure we can find five people who meet this criterion.

The competition for money will increase. Evaluations will be asked to play an increasingly potent role in decision-making.

This dilemma presents itself. The kind of information policy makers need--or feel they need is not of this descriptive nature, which has typically characterized an evaluation, but rather one which

can yield inferences about and contrasts between programs. I see a movement clearly underway towards a "hyper-evaluation" which transcends the classical evaluation design and is more akin to an experimental or research design.

I believe this maturation of effort will yield usable results. It will certainly demand a rigor beyond most of our present efforts.